

## Vinculación profesional para el monitoreo de servicios de TICs utilizando Big Data.

Marcelo Dante Caiafa<sup>1</sup>

Ariel Aurelio<sup>1</sup>

Adrián Marcelo Busto<sup>1</sup>

<sup>1</sup>*Departamento de Ingeniería e Investigaciones Tecnológicas,  
Universidad Nacional de la Matanza, Argentina*

[mcaiafa@unlam.edu.ar](mailto:mcaiafa@unlam.edu.ar), [aaurelio@unlam.edu.ar](mailto:aaurelio@unlam.edu.ar), [abusto@unlam.edu.ar](mailto:abusto@unlam.edu.ar)

### Resumen

La articulación entre la formación universitaria y la profesional es uno de los objetivos de nuestra labor educativa. El interés del trabajo es poner de manifiesto las competencias relevantes del estudiante en un proceso concreto de vinculación entre el ambiente académico y el ambiente productivo.

La investigación se basa en el desarrollo de herramientas para el monitoreo de servicios de los sistemas TIC (tecnología de la información y la comunicación). En esta articulación universidad-empresa se utiliza tecnología Big Data para el procesamiento de un gran volumen de datos contenido en los archivos generados por los sistemas de comunicaciones de una organización de servicios masivos que atiende a más de dos millones de clientes.

Se pretende construir un tablero de control basado en tecnología de código abierto ELK (Elasticsearch-Logstash-Kibana), que se abastecerá con los registros de detalle de llamada CDRs (Call Detail Records) generados por la plataforma de comunicaciones.

El objetivo es validar si el dominio de competencias técnicas es condición suficiente o si las habilidades blandas también son necesarias para lograr un eficiente desempeño en un proyecto de estas características.

**Palabras Clave:** Big Data, Tecnología de la Información, Competencias Profesionales del ingeniero del sector TIC, ELK.

### Introducción a Big Data

La generación de datos se ha incrementado tanto que su tratamiento con los sistemas tradicionales se complica. Este crecimiento hace que requiera de herramientas adecuadas para la comprensión de la información generada. De allí que el término Big Data es de los más abordados en la industria de las tecnologías de la información [1].

Originalmente ha sido típicamente empleado para referirse a la solución al crecimiento exponencial de los datos para su almacenamiento, procesamiento y análisis [2]. Se enumeran algunos de sus beneficios [3]:

- “Optimización del cálculo y la precisión algorítmica para reunir, analizar, enlazar y comparar conjuntos de grandes cantidades de datos”
- “Identificación de patrones para la toma de decisiones en distintos ámbitos (económico, social, técnico, legal, etc)”

Según otros autores, la definición de Big Data es un proceso de exploración, desarrollo y aplicación de algoritmos escalables, infraestructuras y herramientas para organizar, integrar, analizar y visualizar, grandes cantidades de datos, complejos y heterogéneos [4]. Esta heterogeneidad se debe tanto a su volumen como a la variedad de fuentes. Algunos autores estiman que la producción de datos se dobla cada 40 meses [5]. Otros atributos importantes de los datos son la velocidad, su veracidad y su variedad [6].

- Velocidad: esto hace referencia tanto en su generación como en su análisis.
- Veracidad: cualquier actividad puede generar datos. Deben ser fiables, íntegros y auténticos.
- Variedad: hace referencia a los distintos formatos y tipologías de datos.

La herramienta seleccionada para el desarrollo del presente trabajo tiene como motor de búsqueda Elasticsearch y de acuerdo a la clasificación de plataformas Big Data de Matt Turck [7] cae en categoría de big data analítica.

Los tipos de analítica se clasifican [8]:

- a) Analítica descriptiva: ilustra los datos de los resultados recopilados durante un intervalo de tiempo.
- b) Analítica de diagnóstico: busca la causa raíz de un problema.
- c) Analítica predictiva: utiliza datos pasados para realizar pronósticos.
- d) Analítica prescriptiva: está dedicada a encontrar la solución más adecuada.

Por su diversidad los datos se agrupan en tres tipologías: estructurados, semiestructurados y no estructurados. Los primeros tienen un formato específico, están etiquetados y resultan de fácil acceso. Los segundos cuentan con algún tipo de estructura, generalmente en formatos XML o HTML con etiquetas de texto y meta data. Los últimos no disponen almacenamiento estructurado [9].

La aplicación del Big Data en la generación de información referidos al análisis de grandes conjuntos de datos para ayudar en la toma de decisiones presenta múltiples ejemplos [10]. A modo de ejemplo se enumeran algunos:

- Óptimo almacenamiento de datos masivos.
- Integración de datos que provienen de distintos sistemas
- Análisis de la información para facilitar la detección de fallos.

A modo de ejemplo se indican algunos casos de uso basado en análisis de datos contenidos en registros CDR [11]:

- Análisis en tiempo real (monitoreo dinámico del tráfico de comunicaciones).
- Eficiencia operativa (disponibilidad de canales, optimización de operación de servicios).
- Mejora en la experiencia del cliente (monitoreo de calidad de servicios)

El perfil del ingeniero está conformado por competencias técnicas y competencias genéricas [16]. Para los profesionales que trabajan en el sector de las TICs, el análisis y la recopilación de datos genera oportunidades para ejecutar procesos de analítica, medición y diagnóstico de incidentes, etc. Estas nuevas competencias técnicas definen un nuevo especialista, conocido genéricamente como “científico de datos”. Este perfil requiere habilidades en el manejo de información y la mejora de la calidad y las relaciones entre conjuntos de datos [12]. Las principales se centran en el diseño de base de datos, desarrollo de procesos de extracción, transformación y carga de datos de múltiples fuentes. Este proceso conocido por sus siglas en inglés, ETL (extract, transform and load) [13].

Según Pratt, las empresas "requieren personal para interpretar y construir tableros de visualización, de modo que la mesa ejecutiva pueda comprender la información que contienen [14]. Otros autores [15] indican que entre los nuevos perfiles requeridos por el mercado laboral están:

1. Analista de datos: ejecuta análisis estadísticos con distintas fuentes de información.
2. Científico de Datos: aplica matemáticas, estadística y programación (python) para explorar y examinar datos que provienen de fuentes variadas y desconectadas.
3. Gerente de datos (Chief Data Officer): un ejecutivo que actúa como responsable final de la gestión de datos y sus políticas de privacidad.

Una encuesta realizada a 683 directivos de américa latina [17], evidenció que el 69% de

los encuestados no utilizan la data almacenada para la toma de decisiones, sólo un 12% logra aprovecharlos, como se ve en figura 1.

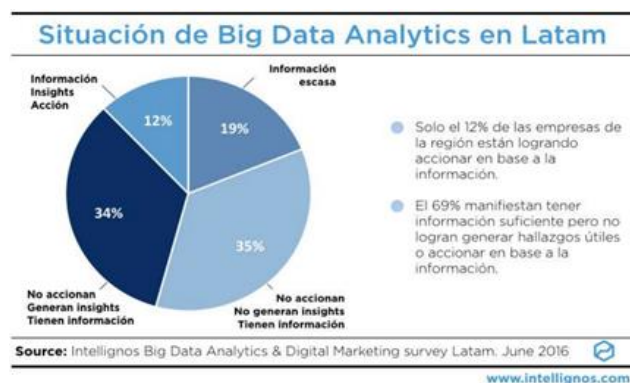


Figura 1: Situación de Big Data Analytics en Latam  
Fuente: Intellignos, 2016.

La consultora internacional ICD (International Data Corporation), espera para este año 2021 un incremento del 129% de gasto en *Big Data Analítica* en América Latina. Las industrias que más están cambiando son las que tienen mayor contacto con usuarios finales, es decir de consumo masivo, como venta minorista, banca, salud, manufactura y logística [18].

En las últimas décadas Argentina ha desarrollado el sector de Software y Servicios Informáticos, con su gran dinámica permitiría aprovechar estos avances en el crecimiento y desarrollo del país [19]. Según lo expresado por algunos autores, las empresas fomentan la creación de nuevas estrategias para la toma de decisiones. De allí que Big data se convierta en uno de los principales aspectos a considerar en los ámbitos comercial, científico y social, debido al gran impacto económico e innovador que representa [20].

Según la Cámara de Empresas de Software y Servicios Informáticos, entre 2003 y 2012 la facturación del mercado de TIC en la Argentina creció un 19,8% anual en promedio. El país cuenta con oportunidades para generar una plataforma para la industria de Big Data

mediante la promoción de políticas adecuadas. Sin embargo, los resultados de la Encuesta Big Data 2014, muestra una demanda escasa explicada en parte por la falta de conocimiento y limitaciones en la infraestructura. De acuerdo al relevamiento del CIECTI (Centro interdisciplinario de estudios en ciencia, tecnología e innovación) las instituciones científicas revelan una desconexión entre la producción académica y los desarrollos de empresas locales [21].

## Problemática a investigar

El presente trabajo, basado en la necesidad de consolidación de datos para aportar valor a la gestión de servicios de tecnología, pretende dar respuesta a la desconexión citada en el apartado anterior.

El estudio se enfoca en la articulación entre el mundo académico y el mundo profesional. Para ello estudia el proceso de construcción de una herramienta de gestión de los servicios de TICs basada en tecnología Big Data.

Algunos autores como Garousi [22] dentro de su propuesta de enseñanza-aprendizaje incorporan proyectos reales, estrategias para desarrollar habilidades y métodos de evaluación en la enseñanza de ingeniería de software basada en la ejecución de proyectos universidad–empresa.

Los objetivos de la investigación apuntan a conocer sobre las habilidades puestas en práctica por el profesional que se desempeña en un proyecto de estas características. De allí surge la pregunta de investigación:

¿Al momento del abordaje de un proyecto que vincula Big Data con los servicios de TICs, es suficiente con el dominio de competencias técnicas específicas, o es preciso además el dominio de competencias genéricas?

Los objetivos del estudio son:

1. Identificar los tipos de competencias profesionales requeridas para lograr un eficiente desempeño en un proyecto de estas características.
2. Construir los indicadores, mediante la selección de los distintos parámetros a partir de los datos disponibles, de modo que reflejen el tráfico telefónico desagregado por categorías.
3. Elaborar una herramienta para el monitoreo del servicio de atención al cliente mediante un tablero de control que consolide información de los distintos canales.

Se pretende aportar valor a la formación del perfil del profesional que trabaja en el sector TIC poniendo en relevancia las habilidades necesarias para lograr un eficiente desempeño en la explotación de datos de los sistemas de una empresa del mercado local.

Como indican algunos autores, las TICs son relevantes al momento de considerar esfuerzos de analítica que procese datos masivos. Este tipo de proyectos no son sólo de implementación de infraestructura tecnológica, por su naturaleza y potencial impacto son distintos, son estratégicos para sustentar la toma de decisiones basada en evidencia [23].

Por un lado, las tareas de integración de plataformas tecnológicas demandan un dominio de competencias técnicas o habilidades duras sobre cada una de ellas, es decir tanto de la tecnología de Big Data como de los sistemas de TICs.

Por otro lado, algunos autores que consideran el ambiente interno de las empresas indican que la relación entre tecnología y habilidades blandas en entornos de big data influyen positivamente en la adopción de nuevas tecnologías [24].

Este trabajo hará foco en los servicios de comunicaciones como la telefonía, video y colaboración. Los datos se obtienen de los registros generados por servidores de comunicaciones de infraestructura Cisco CUCM (Cisco Unified Communication Manager).

**Top 4 Enterprise PBX and UC Voice Equipment Vendors by Global Revenue Share in 3Q14**



© Infonetics Research, Enterprise Unified Communications and Voice Equipment: Quarterly Market Share, Size, and Forecasts: November 2014

**Figura 2: Proveedores de plataformas UC corporativas**

Fuente: Infonetics Research, 2014.

La figura 2 detalla la distribución de mercado de los distintos proveedores de plataformas de comunicaciones unificadas, conocida por sus siglas en inglés como UC (unified communications).

## Plataforma Tecnológica

La plataforma tecnológica de Big Data seleccionada para este trabajo es el conjunto ELK (Elasticsearch-Logstash-Kibana) dada su versatilidad y sus características.

Elasticsearch actúa como repositorio de información que almacena los documentos que indexa. No requiere de un esquema predefinido, ya que la misma colección de documentos puede contener una estructura distinta. Esto lo hace flexible y escalable para manipular datos de gran volumen.

Una de sus ventajas es la capacidad de responder a consultas en tiempo real. Se puede

desplegar sobre un conjunto de máquinas que aporta redundancia y comparte el trabajo de las consultas, aunque en este caso se utilizó una estructura de nodo único.



Figura 3: Plataforma tecnológica empleada  
Fuente: Elasticsearch B.V

Es una solución conformada por tres componentes de donde deriva su acrónimo ELK. (Elasticsearch, Logstash y Kibana), como se ve en la figura 3.

Logstash: es el pre-procesamiento, recoge los datos y los procesa antes de almacenarlos en sus bases. Requiere Java Virtual Machine y corre sobre distintos sistemas operativos.

Elasticsearch: es una base de datos distribuida. Distribuye información y procesamiento en todos los nodos y tiene alta disponibilidad.

Kibana: es la herramienta de presentación, donde se genera la visualización de la información con filtros y tableros de comando. Los tres componentes definen un conjunto de herramientas de código abierto que se combinan para crear una herramienta de administración de registros permitiendo la monitorización, consolidación y análisis de archivos de texto en múltiples servidores. Permite solucionar tres inconvenientes: falta de consistencia (múltiples dispositivos con formatos distintos); formato de tiempo (cada log puede disponer de una referencia temporal distinta), descentralización (cuando los archivos están distribuidos en múltiples rutas). Elasticsearch es un motor de búsqueda basado en Lucene. Dispone de una interfaz de programación de aplicaciones API (Application Programmable Interface) de código abierto para recuperación de información distribuida bajo la licencia de software de Apache desarrollada en Java. Provee un motor de búsqueda de texto completo, distribuido y con capacidad de procesamiento paralelo con

una interfaz web RESTful. Su interfaz web HTTP y documentos JSON, permiten interactuar de forma sencilla con su núcleo y realizar búsquedas de texto completo.

Entre sus características se destacan el procesamiento de datos en tiempo real, funciones de búsqueda con diversos complementos que hace más amigable el análisis y sistema distribuido para procesamiento en paralelo.

Uno de los motivos que llevaron a su adopción son: funcionalidades con bajo costo técnico, disponibilidad de optimización y escalabilidad.

Entre sus desventajas se puede citar la dificultad para la configuración del Logstash en la indexación de datos que sólo soporta documentos de formato JSON. Esto la limita ya que no admite CSV o XML.

El presente trabajo utilizó ELK versión 7.10.0 corriendo Windows 10 JVM versión 8.0 sobre un nodo único.

## Desarrollo del trabajo

El desarrollo del trabajo se estructura en cinco instancias de ejecución secuencial.

En la primera etapa se realizaron las tareas de relevamiento de infraestructura de red, el detalle del direccionamiento IP y la arquitectura del sistema de comunicaciones para conocer la fuente de datos.

En la segunda etapa se obtienen los registros CDR, se identifican los parámetros relevantes y se elabora un diccionario de datos para la correcta interpretación de los datos.

La tercera etapa fue la indexación de la base de datos. A cada campo seleccionado del CDR se le asigna un tipo de parámetro específico para que el motor de búsqueda pueda procesarlo pasando de archivo de texto a documento Json. La etapa cuatro ejecuta el proceso ETL. Se elabora un archivo de texto a partir del cual se configura el módulo Logstash para la ingesta de datos dentro de la base correspondiente.

En la última etapa se realizan las búsquedas de acuerdo a las variables a medir de forma que con Kibana pueda consolidar los indicadores en un tablero de control para su presentación.



## Contextualización de la fuente de datos.

Para alcanzar los objetivos propuestos se utilizó como fuente de datos registros CDRs, generados por los sistemas de comunicaciones de una organización con más de dos millones de clientes dedicada al servicio masivo. Está compuesta por miles empleados con cientos de sucursales a lo largo del territorio nacional. Dado la extensa normativa que regula sus productos, su estrategia de competencia se focaliza en la diferenciación a partir de la calidad de atención al cliente, lo que potencia el valor del presente trabajo.

Dentro del proceso de vinculación fueron necesarias distintas sesiones de trabajo entre ambos equipos de trabajo, del lado de la universidad y del lado de la empresa. Las mismas se enfocaron en descubrir detalles de la infraestructura y aspectos funcionales de los servicios referidos entre los que se citan:

- la arquitectura funcional de los sistemas que son fuentes de datos,
- el plan de direccionamiento IP de los segmentos de red LAN, MAN y WAN,
- el plan de numeración del servicio telefónico y su interconexión con prestadores,
- la estructura de los diferentes modelos de atención al cliente a estudiar.

Estas actividades fueron clave para la adecuada contextualización de los datos y su correspondiente interpretación.

## Diccionario de datos

El diccionario de datos sirve para dar significado a cada campo del registro.

Table 1. Ejemplos de parámetros en CDR

<i>Parámetro</i>	<i>Significado</i>
CallingPartyNumber	Call origin number
CalledPartyNumber	Call destination number
Duration	Duration Time (sec)
lastRedirectDn	Last transferred number
origDeviceName	Id source device
destDeviceName	Id target device
origIpv4v6	Source IP address
destIpv4v6	Destination IP address

Estos archivos se generan cada vez que se realiza o recibe una llamada. Su ciclo de vida inicia con la generación de una llamada y se actualiza con los eventos que ocurren durante la misma (duración, final, transferencia, etc). Los registros presentan distintos formatos acordes a la tecnología empleada por el operador, como se muestra en la tabla 1.

Estos datos pueden ser utilizados para los procesos de carga, liquidación, facturación, eficiencia de la red, detección de fraude, servicios de valor agregado, inteligencia de negocios, etc. Adicionalmente contribuyen a mejorar servicios y procesos existentes en distintas áreas [25].

## Indexación de la base de datos

Para generar la base de datos es preciso crear un índice en Kibana. Se definen los campos necesarios para dar formato a la estructura de datos que se espera recibir. Esto se ejecuta en el apartado DevTools con una petición PUT. La figura 4 muestra parte del archivo utilizado:

```
PUT /cdr2020DBv2 {
  "mappings": {
    "properties": {
      "cdrRecordType":{"type":"integer"},
      "globalCallID_callManagerId":{"type":"integer"},
      "globalCallID_callId":{"type":"integer"},
      "origLegCallIdentifier":{"type":"integer"},
      "dateTimeOrigination":{"type":"integer"},
      /se utiliza tipo entero porque de acuerdo al
      diccionario de datos la fecha y hora están en formato
      UNIX (cantidad de segundos transcurridos desde 01/01/1970)
      "dateTimeOrigination_formatted":{"type":"date"},
      /se realiza la transformación del campo a partir
      del tipo de archivo .conf del proceso ETL.
      "dateTimeConnect_formatted":{"type":"date"}
      "origNodeId":{"type":"integer"},
      "origSpan":{"type":"integer"},
      "origIdAddr":{"type":"integer"},
      "origIpv4v6Addr":{"type":"ip"},
      /se utiliza tipo IP, en lugar del tipo entero ya
      que Kibana permite discriminar rangos de IP.
      "destIpv4v6Addr":{"type":"ip"},
      "callingPartyNumber":{"type":"integer"},
      "callingPartyNumberPartition":{"type":"text"},
      /se utiliza tipo texto porque es el formato en que
      se espera recibir el nombre origen de llamada.
      "origNodeId":{"type":"integer"},
      "callingPartyUnicodeLoginUserID":{"type":"text"},
      "origCause_location":{"type":"integer"},
      "dateTimeImport":{"type":"date"},
      "origCause_value":{"type":"integer"},
      "origMediaPrecedenceLevel":{"type":"integer"},
      "origMediaTransportAddress_IP":{"type":"integer"},
    }
  }
}
```

Figura 4: Indexación de la base de datos

Fuente: Elaboración propia

## Proceso ETL

El proceso ETL (Extract, Transform and Load) es el proceso de toma de datos, la adaptación de sus campos y tipos, y la carga de datos a la base. Esto se realiza configurando el logstash con el archivo que se ilustra en la figura 5. Se identifican claramente tres instancias. La primera es la entrada (input), que consiste en indicar la ruta del archivo de donde se extraen los datos. El siguiente paso es el filtro (filter) donde se listan todos los campos que contendrá en formato separado por comas. En el caso particular de las fechas, se pueden convertir los datos que se reciben en formato UNIX a un tipo dato. Por último, se tiene la salida (output) que indica el nombre del índice donde quedarán cargados los datos. El agregado del comando "stdout" habilita la opción para observar la carga en pantalla.

```
input {file {
  path => "C:/Users/unlam/CDR2020.txt"
  start_position => "beginning"
}}
filter {csv {columns => ["cdrRecordType",
  "globalCallID_callManagerId",
  "origLegCallIdentifier",
  "dateTimeOrigination",".....",
  "origVideoCap_resolution_Channel2"]}
  date {match => ["dateTimeOrigination","UNIX"]
  target => ["dateTimeOrigination_formatted"]
  date {match => ["dateTimeConnect","UNIX"]
  target => ["dateTimeConnect_formatted"]
  date {match => ["dateTimeOrigination","UNIX"]
  target => ["dateTimeOrigination_formatted"]
  date {match => ["dateTimeConnect","UNIX"]
  target => ["dateTimeConnect_formatted"] }
output {stdout {}
  elasticsearch {index => "cdr2020DBv2"} }
```

Figura 5: Proceso ETL  
Fuente: Elaboración propia

## Definición de indicadores

A partir de las recomendaciones de la Unión Internacional de Telecomunicaciones [26] se definen los indicadores más relevantes para la medición de las variables:

International outgoing traffic: Tráfico saliente a destino internacional

Outgoing Mobile Traffic: Tráfico saliente a celulares

Outgoing Long-Distance Traffic: Tráfico saliente a destino Nacional

Donde el tráfico es la cantidad de llamadas por unidad de tiempo.

Duración: es la cantidad de tiempo que transcurre durante la interacción total.

## Resultados Obtenidos

El parámetro finalCalledPartyNumber se utilizó para clasificar el tráfico de voz según las categorías destino: Locales, Nacionales, Emergencia, Internacionales y Celulares. Esto permite registrar la distribución del tráfico telefónico y el costo operativo del servicio de los proveedores de la PSTN. El filtro se aplicó a diferentes periodos para evidenciar los cambios provocados por las restricciones generadas por la pandemia. La figura 6 compara los datos correspondientes a los periodos prepandemia (Nov/Dic 2019) y período durante la pandemia (Nov/Dic 2020)

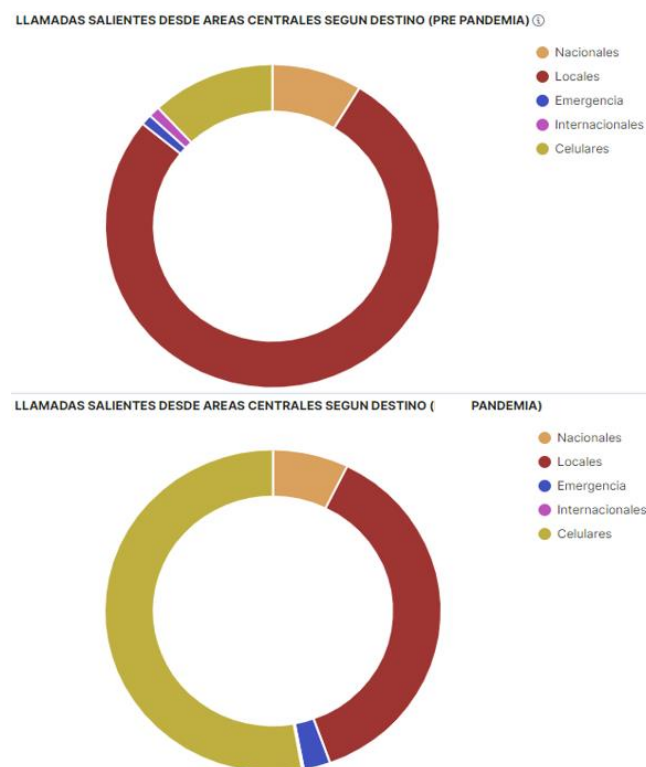


Figura 6: Distribución del tráfico saliente  
Fuente: Elaboración propia

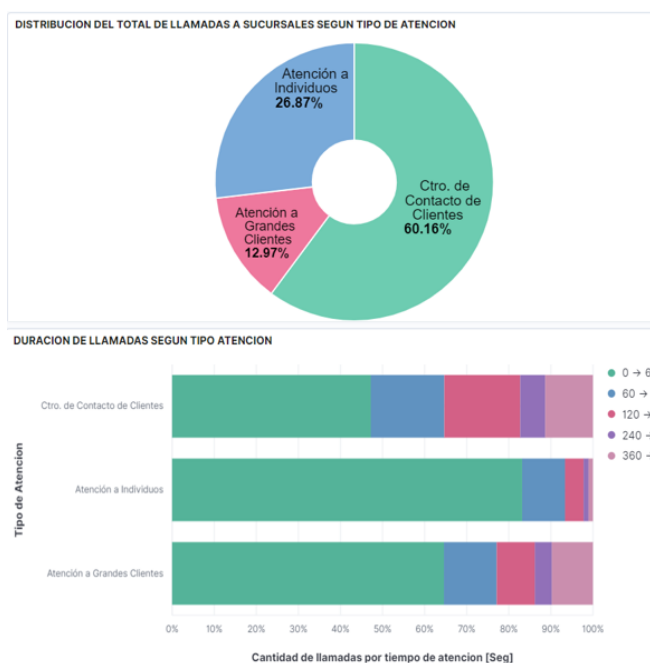


Figura 7: Indicadores de atención al cliente  
Fuente: Elaboración propia

Se utilizaron 2 gráficos para construir el tablero que mide el nivel de calidad de atención al cliente y se muestra en la figura 7. El gráfico superior tiene la distribución de llamadas entrantes clasificadas por tipo de atención. La atención se conforma en tres grupos: centro contacto con clientes, atención a individuos y grandes clientes. El gráfico inferior detalla la duración de las llamadas en grupos segmentados por minuto.

Respecto a las habilidades genéricas transversales puestas de manifiesto, fundamentalmente en las etapas 1, 2 y 5 del proyecto se relevaron la capacidad de indagación, pensamiento crítico, organización del tiempo y comunicación efectiva.

## Conclusiones

El trabajo permitió valorar la experiencia de enseñanza de la ingeniería a través de la participación de estudiantes en un proyecto concretos de articulación academia-empresa, donde la industria aportó contexto productivo y especialistas técnicos y de negocio.

Respecto a los datos de tráfico del servicio de telefonía, comparando los mismos meses de 2019 (preCovid-19) frente al último 2020 (durante Covid-19) se observa un incremento del 320% para la categoría destino Celulares y una reducción del 50% de la categoría destino Locales. Esto refleja la forma en que afectaron los cambios impuestos por las restricciones de la pandemia, que redujo la concurrencia de personal a las sucursales a cambio de utilizar teléfonos celulares.

En el tablero de atención del cliente se observa que los llamados atendidos por el centro de contacto (CC) representan un 60% del total de llamadas a sucursales. El tráfico restante se distribuye en dos tercios para el sector individuos y uno a grandes empresas.

En el CC el 45% de las llamadas duran menos de 1 minuto. El personal del CC adapta su disponibilidad de recursos a niveles de productividad predefinidos.

Comparando las llamadas respondidas por oficiales en sucursal con duración mayor a 5 minutos, se observa que grandes empresas tienen un 10% del total, mientras que individuos es sólo 1%.

En este proceso de vinculación con el ambiente productivo se comprobó que, si bien requirió de competencias técnicas, puso de manifiesto la necesidad de competencias genéricas para las tareas de interpretación de los datos. Se comprobó capacidad de indagación para adecuada interpretación de la fuente de datos y de la arquitectura de servicios.

Desde el punto de vista de los estudiantes, el análisis de esta experiencia permitió identificar como ventaja, la aplicación de los conocimientos técnicos en ambientes profesionales y el reconocimiento de las habilidades blandas como factores críticos de éxito dentro del proyecto.

Estas últimas resultaron clave particularmente en la etapa de contextualización de datos.

Como futuros trabajos se propone el estudio de metodologías de evaluación de competencias en proyectos de vinculación. Esto constituye un insumo importante para la valoración del desarrollo de habilidades blandas en entornos de integración de plataformas tecnológicas.



## Bibliografía

- [1] K. Regunath. *Real-Time intrusion detection system for Big Data*. International Journal of Peer to Peer Networks (IJP2P), vol. VIII, n° 1, 2017. Disponible en <https://link.springer.com/article/10.1007/s11227-015-1615-5>
- [2] J. Manyika. *Big Data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011.
- [3] J. Routledge. *Critical questions for Big data*. Cambridge: Danah Boyd & Kate Crawford. 2013. Disponible en <http://www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878>
- [4] S. Mutula. *Big Data Industry: Implication for the library and information sciences*. 2016. African Journal of Library Archives and Information Science, 26(2), 93–96. Disponible en: <https://www.ajol.info/index.php/ajla/article/view/167425>
- [5] A. McAfee y E. Brynjolfsson, “*Big Data: the management revolution*”, 2012. Editorial HBR, Harvard Business Review, vol. 90, n°10, p. 61- 68.
- [6] Y. Demchenko. “*Defining the Big Data Architecture Framework*”. 2013. University of Amsterdam, disponible en: [http://bigdatawg.nist.gov/uploadfiles/M0055v1\\_7606723276.pdf](http://bigdatawg.nist.gov/uploadfiles/M0055v1_7606723276.pdf)
- [7] M. Turck. *Firing on All Cylinders: The 2017 Big Data Landscape*. 2020. Disponible en <http://mattturck.com/bigdata2020/>
- [8] Y. Riahi, & S. Riahi. “*Big Data and Big Data Analytics: Concepts, Types and Technologies*”. International Journal of Research and Engineering, IJRE | Vol. 5 No. 9, pág 525. 2018
- [9] S. Soares, 2012. *Big Data Matchmaker*. Five Data Types You Need To Explore Today. Disponible en <https://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/>
- [10] M. Machado. *Applications of Big Data for Development*. 2016. Disponible en <https://www.engineeringforchange.org/news/applications-of-big-data-for-development/>
- [11] Elagib, Hashim & Olanrewaju. “*CDR Analysis using Big Data Technology*”. 2016. Publicado por International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering. Disponible en <https://ieeexplore.ieee.org/document/7381414>
- [12] J. Morato, S., Sanchez- Cuadrado, & Fernández Bajón. 2016. *Trends in the technological profile of information professionals*. El Profesional de La Información, 25(2), 169-178. Disponible en: <https://doi.org/10.3145/epi.2016.mar.03>
- [13] C. Dobre y F. Xhafa. *Parallel programming paradigms and frameworks in Big Data Era*. 2014. International Journal of Parallel Programming. 42(5), 710-738. <https://doi.org/10.1007/s10766-013-0272-7>
- [14] M. Pratt. “*Hottest Tech Skills for 2016*”. Computer World, (diciembre 2015), 29–34. Disponible en <http://bit.ly/2oPyDeD>
- [15] J. Monroy Osorio y Marian Villa. “*Uso de las plataformas de big data como herramienta en inteligencia de negocio*”. 2017. 5-6. Disponible : [https://www.researchgate.net/publication/340173523\\_Uso\\_de\\_plataformas\\_de\\_Big\\_Data\\_como\\_herramienta\\_en\\_inteligencia\\_de\\_negocio](https://www.researchgate.net/publication/340173523_Uso_de_plataformas_de_Big_Data_como_herramienta_en_inteligencia_de_negocio)
- [16] Carrer Space. *Perfiles de capacidades profesionales genéricas de TIC. Capacidades profesionales futuras para el mundo del*

mañana. Oficina de Publicaciones Oficiales de las Comunidades Europeas. Luxemburgo, 2001. Disponible en <http://www.carrer-space.com>

[17] Intellignos. *Situación de Big Data Analytics en Latam*. 2016. Intellignos Big Data Analytics & Digital Marketing en Latinoamérica. Junio 2016.

[18] IDC. “En América Latina habrá un incremento del 129% en gasto de Big Data y analítica”. 2018. Entrevista a Oliver Aguilar, Disponible en <https://inversorlatam.com/big-data-habra-un-129-de-aumento-de-gasto-en-america-latina/>

[19] F. Barletta, M. Pereira, V. Robert y G. Yoguel (2013), “Argentina: dinámica reciente del sector de software y servicios informáticos”. Revista cepal, N° 110, agosto. Disponible en <https://repositorio.cepal.org/handle/11362/11618>

[20] R. Barranco: “¿Qué es big data?”. IBM, México [citado 15 marzo, 2021]. Disponible en: <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

[21] F. Malvicino y G. Yoguel. “Big Data. Avances recientes a nivel internacional y perspectivas para el desarrollo local. 2016. CIECTI (Centro interdisciplinario de estudios en ciencia, tecnología e innovación). Disponible en <http://www.ciecti.org.ar/wpcontent/uploads/2017/07/DT3-BigData-avances-y-perspectivas-de-desarrollo-local.pdf>

[22] V. Garousi. Incorporating real-world industrial testing projects in software testing courses: opportunities, challenges, and lessons learned. En Software Engineering Education and Training (CSEE&T 2011), 24th IEEE-CS Conference on, pp. 396-400. IEEE.

[23] P. Rodríguez, N. Palomino y J. Mondaca. *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. BID. Sector de Conocimiento y Aprendizaje. 2017

[24] Caputo, F., Cillo, V., Candelo, E., & Liu, Y. (2019). Innovating through digital revolution: The role of soft skills and Big Data in increasing firm performance. *Management Decision*, 57(8), 2032–2051.

[25] D. Agrawal, P. Bernstein, E. Bertino, S. Davidson y U. Dayal, "Challenges and Opportunities with Big Data". USA, CYBER CENTER TECHNICAL REPORTS, 2011, <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1000&context=cctech>

[26] ITU. List of indicators included in the World Telecommunication/ICT Indicators database, December 2016. International Union of Telecommunications. Disponible en [https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2016/WTID\\_2016\\_ListOfIndicators.pdf](https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2016/WTID_2016_ListOfIndicators.pdf)